

1. [Suggested Plan for Teaching the Course](#)
2. [Sampling and Data: Teacher's Guide](#)
3. [Descriptive Statistics: Teacher's Guide](#)
4. [Probability Topics: Teacher's Guide](#)
5. [Ch. 4: Discrete Random Variables](#)
6. [Continuous Random Variables: Teacher's Guide](#)
7. [Normal Distribution: Teacher's Guide](#)
8. [Central Limit Theorem: Teacher's Guide](#)
9. [Confidence Intervals: Teacher's Guide](#)
10. [Hypothesis Testing: Single Mean and Single Proportion: Teacher's Guide](#)
11. [Hypothesis Testing: Two Population Means and Two Population Proportions: Teacher's Guide](#)
12. [Ch 11: The Chi-Square Distribution](#)
13. [Ch 12: Linear Regression and Correlation](#)
14. [Ch 13: F Distribution and ANOVA](#)

Suggested Plan for Teaching the Course

This module provides a suggested plan for teaching a statistics course using the Collaborative Statistics collection (col10522).

Each chapter is **interactive**. Students should fill in the blanks and answer the questions.

At the end of each chapter is at least one **practice**. The practice leads the students step-by-step through problems. We, the authors, start the practices in class with students working in groups of 2, 3, or 4. The students finish the practices at home. The **practice** is after the chapter reading but before the homework.

The back of the book contains answers to the odd-numbered **homework** problems. **In this plan (this document), the suggested homework is listed at the end of the chapter discussion.**

At the end of each chapter (after the homework), there is at least one **lab**. The labs use real data collected by the instructor or the students or both. We often use the class to collect data. Labs may be done in groups and are an excellent teaching tool especially if they are started in class. The book contains the following labs:

- **Ch. 1**Data Collection Lab I (number of movies viewed)
- **Ch. 1**Sampling Experiment Lab II (table of restaurants provided)
- **Ch. 2**Descriptive Statistics Lab (number of pairs of shoes)
- **Ch. 3**Probability Lab (counting M&M's)
- **Ch. 4**Discrete Distribution Lab I (picking playing cards)
- **Ch. 4**Discrete Distribution Lab II (Tet game)
- **Ch. 5**Continuous Distribution Lab (generate random numbers)
- **Ch. 6**Normal Distribution Lab I (Terry Vogel's lap times provided)
- **Ch. 6**Normal Distribution Lab II (measure pinkie fingers)
- **Ch. 7**Central Limit Theorem Lab I (counting change)
- **Ch. 7**Central Limit Theorem Lab II (cookie recipes)
- **Ch. 8**Confidence Interval Lab I (real estate prices)
- **Ch. 8**Confidence Interval Lab II (students born in state)
- **Ch. 8**Confidence Interval Lab III (heights of women)

- **Ch. 9**Hypothesis Testing Lab - Single Mean and Single Proportion (3 tests)
- **Ch. 10**Hypothesis Testing Lab - Two Means and Two Proportions (3 tests)
- **Ch. 11**Chi-Square Goodness of Fit Lab I (grocery store receipts)
- **Ch. 11**Chi-Square Test for Independence Lab II (favorite snack/gender)
- **Ch. 12**Regression Lab I (distance from school vs. cost of supplies this term)
- **Ch. 12**Regression Lab II (number of pages in textbook vs. cost of textbook)
- **Ch. 12**Regression Lab II (weights vs. fuel efficiency)
- **Ch. 13**ANOVA Lab (fruits, vegetables, breads)

Because the authors use technology heavily in the course (making many class periods a lab), **we typically choose to do 6 labs during the quarter.** The labs are best done in groups of 2, 3, or 4.

There are five **projects** in the book. The Univariate Data project covers the ideas in chapters 1 and 2. The Continuous Distributions and Central Limit Theorem project covers idea in chapters 5, 6, and 7. The Hypothesis Testing - Article and the Hypothesis Testing - Word project covers ideas in chapters 8 and 9. The Bivariate Data, Linear Regression and Univariate project covers ideas in chapters 1, 2, and 12. Projects are done in groups of 2, 3, or 4.

There are **Practice Finals** with answers and **Data Sets** in the text. One of the Chapter 6 Labs uses one of the data sets. Going over the Table of Contents for this collection with the students is recommended.

We carry probabilities to 4 decimal places.

The number of days (a "day" is a 50 minute period) based on a quarter system (10 weeks of class, 1 week of finals) it takes to cover a chapter is below. At De Anza, we are on a quarter system. In a semester, you could spend more time analyzing real data. The material is meant to be covered in one quarter or in one semester.

- **Ch. 1** Introduction - 2 days
- **Ch. 2** Descriptive Statistics - 4 days
- **Ch. 3** Probability Topics - 4 days
- **Ch. 4** Discrete Random Variables - 5 days
- **Ch. 5** Continuous Random Variables - 3 days
- **Ch. 6** The Normal Distribution - 3 days
- **Ch. 7** The Central Limit Theorem - 3 days
- **Ch. 8** Confidence Intervals - 4 days
- **Ch. 9** Hypothesis Testing - Single Mean and Single Proportion - 4 days
- **Ch. 10** Hypothesis Testing - Two Means and Two Proportions - 4 days
- **Ch. 11** The Chi-Square Distribution - 4 days
- **Ch. 12** Linear Regression and Correlation - 4 days
- **Ch. 13** Analysis of Variance and F Distribution - 3 days

Sampling and Data: Teacher's Guide

Explain the terms statistics and probability.

Introduce the key terms by an example.

Example:

Students may be interested in the average time (in years) it will take them to earn a B.A. or B.S. Differentiate between population and sample.

Explain data. The book discusses qualitative and quantitative data. Quantitative data is either discrete (countable) or continuous (measurable).

Types of Data

- **Qualitative data** - the city or town a student lives in.
- **Quantitative discrete (countable) data** - the number of T-shirts a student owns.
- **Quantitative continuous (measurable) data** - the amount of time (in hours) a student studies statistics each day.

Sampling

Discuss what a sample is. Stress the importance of sampling randomly and the fact that two random samples from the same population may be different. Doing the two experiments with a fair die (roll the same die 20 times for each experiment and record the frequencies of the faces in the book) will help them understand how samples vary. Using your class as the population, sample 10 men and 10 women. Let the sample be the number of pairs of shoes each student owns. This example illustrates samples which are **not** representative from the same population.

Discuss how to sample data. Though there are numerous ways, the book discusses simple random, stratified, cluster, systematic, and convenience. You may want to discuss other ways of sampling.

Frequency

The last part of the chapter discusses frequency, relative frequency, and cumulative relative frequency. The students should understand how to read the table in the example (heights, to the nearest inch, of male students at ABC College).

Assign Practice

Take some class time and have the students work in groups and complete the [Practice](#).

Assign Homework

Assign [Homework](#) problems: 1 - 17 odds, 19 - 27.

Descriptive Statistics: Teacher's Guide

Graphs are important tools in statistics and probability. Graphs used in this course are the boxplot, the histogram, and the stem-plot. The histogram and boxplot are used extensively while the stem-plot is just demonstrated.

Illustrate Examples

- To illustrate stem-plots, have the students complete Example 2-2 by hand.
- To illustrate histograms, have the students do Example 2-4 by hand and then, if you are using technology, have them do the same example. They can verify their results by looking at the picture.
- Right after Example 2-4, there is an "Optional Collaborative Classroom Exercise" for the students to do that involves the amount of money they have in their pocket or purse.
- To illustrate the boxplot, have the students do Example 2-6. In this example, they will compare two boxplots.

Center of Data

Discuss the measures of "center" - mean (average), median, mode. If you are using technology, it helps to show the students how to use technology to find the measures first. Then do some examples by hand. Distinguish between the symbols used for the sample mean and the population mean. Give an example where the mean is the best measure of the center and a second example where the median is the best example. (Example where median is the better measure: 19, 16, 46, 18, 21. Example where mean is the better measure: 18, 20, 23, 25, 25.) At the end of the chapter, there is a summary of the mean formulas if you desire to go over them.

Spread of Data

Discuss the measures of spread - variance and standard deviation. Stress that the standard deviation is the square root of the variance. Differentiate between the sample and population standard deviations. Dividing by $n - 1$ in the sample variance formula makes the sample standard deviation a better estimator of the population standard deviation. Do one example by hand and have the students participate (the set $\{1, 2, 3\}$ is quick and easy). They will have to calculate the mean first. They should discover how easy it

is to make a numerical error when they calculate standard deviation by hand.

Location of Data

Discuss the measures of location - quartile and percentile. For many students, these measures are difficult. It is better to make up a relative frequency table from an example like the one in the book (the amount of sleep 50 students get per school night) and find quartiles and percentiles. Graphing calculators typically calculate quartiles.

Definition of Value

We introduce the formula

$$\text{Value} = \text{Mean} + (\# \text{ of STDEVs})(\text{Standard Deviation})$$
 in this chapter.

For example, a student with a 74 on the first exam in a statistics class wants to compare his score to a student who received a 70 in another section. If the mean and standard deviation for the first class was 72 and 4, respectively, and the mean and standard deviation for the second class was 68 and 2, respectively, which student did better relative to the class? Solve the equation for #OfSTDEVs in each case.

Assign Practice

Have students work in groups to complete [Practice 1](#) and [Practice 2](#).

Calculator Instructions

If you are using the TI-83 or TI-84 calculator series, go over the calculator instructions in the text for entering data and calculating the sample mean, the sample standard deviation, the quartiles, constructing histograms, and construction boxplots. The calculator instructions can also be found on the Texas Instruments website and the appropriate Guidebook.

Assign Homework

Assign [Homework](#). Suggested problems: 1 - 23 odds, 24 - 30.

Probability Topics: Teacher's Guide

The best way to introduce the terms is through examples. You can introduce the terms experiment, outcome, sample space, event, probability, equally likely, conditional, mutually exclusive events, and independent events AND you can introduce the addition rule, the multiplication rule with the following example: In a box (you cannot see into it), there are 4 red cards numbered 1, 2, 3, 4 and 9 green cards numbered 1, 2, 3, 4, 5, 6, 7, 8, 9. You randomly draw one card (experiment). Let R be the event the card is red. Let G be the event the card is green. Let E be the event the card has an even number on it.

Example:

Event Card Example

1. List all possible outcomes (the sample space). Have students list the sample space in the form $\{R1, R2, R3, R4, G1, G2, G3, G4, G5, G6, G7, G8, G9\}$. Each outcome is equally likely. Plane outcome = $\frac{1}{13}$.
2. Find $P(R)$.
3. Find $P(G)$. G is the complement of R . $P(G) + P(R) = \underline{\hspace{2cm}}$.
4. $P(\text{red card given a that the card has an even number on it}) = P(R | E)$. This is a conditional. Pick the red card out of the even cards. There are 6 even cards.
5. Find $P(R \text{ AND } E)$. (Multiplication Rule:
 $P(R \text{ and } E) = P(E | R)(P(R))$)
6. $P(R \text{ OR } E)$. (Addition Rule:
 $P(R \text{ OR } E) = P(E) + P(R) - P(E \text{ AND } R)$)
7. Are the events R and G mutually exclusive? Why or why not?
8. Are the events G and E independent? Why or why not?

Example:

Exercise:

Problem:

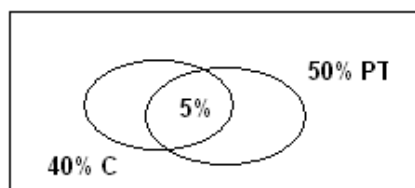
(Optional Topic) A **Venn diagram** is a tool that helps to simplify probability problems. Introduce a Venn diagram using an example.

Example: Suppose 40% of the students at ABC College belong to a club and 50% of the student body work part time. Five percent of the student body works part time and belongs to a club.

Have the students work in groups to draw an appropriate Venn diagram after you have shown them what a Venn diagram basically looks like. The diagram should consist of a rectangle with two overlapping circles. One rectangle represents the students who belong to a club (40%) and the other circle represents those students who work part time (50%). The overlapping part are those students who belong to a club and who work part time (5%).

Find the following:

1. $P(\text{student works part time but does not belong to a club})$
2. $P(\text{student belongs to a club given that the student works part time})$
3. $P(\text{student does not belong to a club})$
4. $P(\text{works part time given that the student belongs to a club})$
5. $P(\text{student belongs to a club or the student works part time})$

Solution:

- **C** student belongs to a club
- **PT** student works part time

Example:**Exercise:**

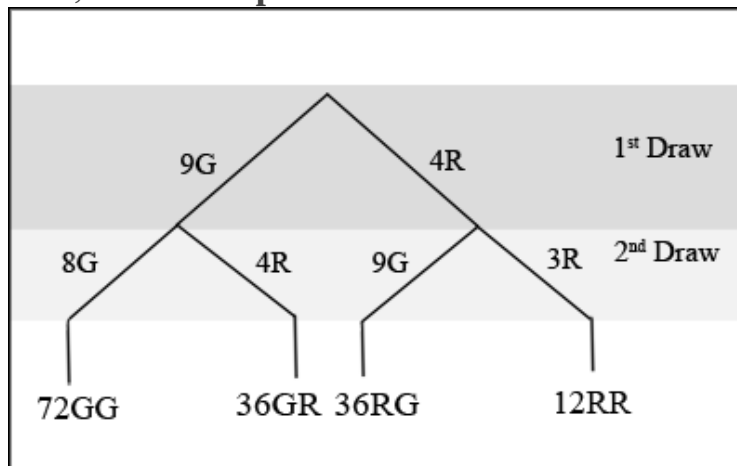
Problem:

Find the following:

1. $P(\text{a child is 9 - 11 years old})$
2. $P(\text{a child prefers regular soccer camp})$
3. $P(\text{a child is 9 - 11 years old and prefers regular soccer camp})$
4. $P(\text{a child is 9 - 11 years old or prefers regular soccer camp})$
- 5.
6. $P(\text{a child is over 14 given that the child prefers micro soccer camp})$
7. $P(\text{a child prefers micro soccer camp given that the child is over 14})$

Tree Diagrams (Optional Topic)

A **tree** is another probability tool. Many probability problems are simplified by a tree diagram. To exemplify this, suppose you want to draw two cards, one at a time, **without replacement** from the box of 4 red cards and 9 green cards.



There are $(13)(12) = 156$ Possible Outcomes. (ex. R1R1, R1R2, R1G3, G3G4, etc.)

Example:

Exercise:

Problem:

Find the following:

1. $P(RR)$
2. $P(RG \text{ or } GR)$
3. $P(\text{at most one } G \text{ in two draws})$
4. $P(G \text{ on the 2nd draw} | R \text{ on the 1st draw})$. The size of the sample space has been reduced to $12 + 36 = 48$.
5. $P(\text{no } R \text{ on the 1st draw})$

Introduce **contingency tables** as another tool to calculate probabilities. Let's suppose an owner of a soccer camp for children keeps information concerning the type of soccer camp the children prefer and their ages. The data is for 572 children.

Type of Soccer Camp Preference	Under 6	6-8	9-11	12-14	Over 14	Row Total
Micro	42	76	46	25	10	199
Regular	8	68	92	105	100	373
Column Total	50	144	138	130	110	572

Assign Practice

Assign [Practice 1](#) and [Practice 2](#) in class. Have students work in groups.

Assign Lab

The Probability Lab is an excellent way to cement many of the ideas of probability. The lab is a group effort (3 - 4 students per group).

Assign Homework

Assign [Homework](#). Suggested problems: 1 - 15 odds, 19, 20, 21, 23, 27, 28 - 30.

Ch. 4: Discrete Random Variables

This module is the complementary teacher's guide for the "Discrete Random Variables" chapter of the Collaborative Statistics collection (col10522) by Barbara Illowsky and Susan Dean.

This chapter introduces **expected value** (long term average) and four of the **common discrete random variables** (binomial, geometric, hypergeometric, and Poisson). The authors cover expected value and two of the discrete random variables (binomial and Poisson). Depending on your background, you may want to cover the binomial (usually required) together with none or some of the other discrete random variables

Random Variables

Explain **random variable** (assigns numerical values to the outcomes of a statistical experiment). Upper case letters denote random variables.

Example: Let X = the number of cars in your household. (The phrase "the number of" tells you that X takes on discrete values.) X takes on the values 0, 1, 2, 3, ...

The Probability Distribution Function

A **probability distribution function (pdf)** is best shown with an example: A controversial drug is given to two patients. Let X = the number of patients cured.

- $P(\text{a cure}) = \frac{5}{6}$
- $P(\text{no cure}) = \frac{1}{6}$

A pdf is easiest to understand in a table.

X	$P(X)$ or $P(X = x)$
0	$P(X = 0) = \left(\frac{1}{6}\right)\left(\frac{1}{6}\right) = \left(\frac{1}{36}\right)$

X	$P(X)$ or $P(X = x)$
1	$P(X = 1) = 2 \left(\frac{1}{6}\right)\left(\frac{5}{6}\right) = \left(\frac{10}{36}\right)$
2	$P(X = 2) = \left(\frac{5}{6}\right)\left(\frac{5}{6}\right) = \left(\frac{25}{36}\right)$

Each probability is between 0 and 1.

The previous example can be used as an example of **expected value** or long term average (μ). Make a third column labeled $(x)(P(x))$. Calculate the three values and add them. The result, $(0)\left(\frac{1}{36}\right) + (1)\left(\frac{10}{36}\right) + (2)\left(\frac{25}{36}\right) = \frac{60}{36} = 1.67$, is the expected number of patients who are cured if the drug is administered many times to two patients.

The binomial is a special discrete pdf or **pattern**. A binomial experiment consists of counting the number of successes in one or more Bernoulli trials. (A Bernoulli trial has only two possible outcomes, success or failure. In every Bernoulli trial, the probability of a success (or failure) remains the same.)

Example:

Exercise:

Problem:

John comes to his stat class and discovers he must take a true-false quiz. There are 20 questions on the quiz. John has not attended class recently and must guess randomly at the questions. Let X = the number of questions John answers correctly out of 20 questions. X takes on the values 0, 1, 2, 3, ..., 20. $P(\text{correct answer: a success}) = 0.5$. John's guessing at the answers is a binomial experiment.

Notation: $X \sim B(20, 0.5)$ where the number of trials, n , is 20 and the probability of a success, p , on any trial is 0.5.

Students can find the mean ($\mu = np$), and the standard deviation ($\sigma = \text{square root of } npq$) either by hand or with technology. (q is the probability of a failure.) Have students help you fill in the blanks and answer the questions:

1. $\sigma =$
2. Draw the graph. (horizontal axis is the number of successes; vertical is the probability of 0 successes, 1 success, 2 successes, ..., 20 successes. Draw vertical lines or boxes.
3.
 - What is the probability that John gets 15 questions correct? $P(X = 15)$
 - More than 15 questions correct? $P(X > 15)$
 - At least 15 questions correct? $(P(X = 15) + P(X > 15))$

A **geometric** experiment takes place when at least one Bernoulli trial is performed and all are failures except the last one which is the only success. Example: Liz likes to play darts. The probability that she hits the bull's eye (success) on any throw is 85%. (Liz is good!) Liz throws darts at the bull's eye **until** she hits it. Let X = the number of times Liz throws the dart at the bull's eye until she hits it. Have students help you fill in the blanks:

Fill in the blanks.

- $X \sim$ _____ ($X \sim G(p)$ where $p = \text{probability of a success} = 0.85$)
- Draw the graph. (Number of throws until the first success versus probability)
- 4. What is the probability that Liz hits the bull's eye for the first time on the third throw? That it takes more than three throws for Liz to hit the bull's eye for the first time? That it takes at least three throws?
- X takes on the values _____.
- $\mu =$ _____. In words, μ is _____

The Geometric Equation

$$P(X = x) = q^{x-1} \cdot p$$

Hypergeometric Distribution

The **hypergeometric distribution** is characterized by choosing a sample without replacement from two distinct groups. One of the two groups is what is of interest in the sample. Some lotteries are based on the hypergeometric distribution. [click to edit note](#)

Example:

Suppose a shipment of 20 tape recorders contains 5 defectives. An inspector randomly chooses 8 of the tape recorders to inspect. He is interested in the number of defectives in the sample of 8. Have the class answer questions similar to those for the binomial and the geometric.

Note: $X \sim H(r, b, n)$ where r = size of the group of interest, b = size of the other group, and n = size of the sample.

Poisson Distribution

The Poisson distribution is concerned with the number of times an event takes place in a certain interval. It is used in the field of reliability. The Poisson approximates the binomial when n is "large" (say, more than 100) and p is "small" (say, less than 0.1).

Example:

Suppose the average number of accidents that occur in a week at a particularly busy intersection is one. The interval is one week. The average is one accident. Let X = the number of accidents that occur in a one week period at the intersection. Have the students help fill in the blanks and answer the questions:

1. $X \sim \underline{\hspace{2cm}}$ ($X \sim P(\mu$ where $\mu =$ one accident))
2. What values does X take on?
3. What is the probability that at most one accident occurs in a week?

The Poisson Distribution Formula

The parameter for the Poisson is the mean, μ . Some books and calculators use the Greek letter, λ (lambda) as the mean. The equation for the Poisson is:

Equation:

$$P(X = x) = \frac{\mu^x \cdot e^{-\mu}}{x!} \quad \text{where } x = 0, 1, 2, 3, \dots$$

Assign Practice

Have the students complete the **portion of the practice** that is appropriate for what you have covered in class. Expected Value, Binomial, and Poisson are dealt with [Practice 1](#), [Practice 2](#), and [Practice 3](#). [Practice 4](#) is based on the Geometric Distribution, while [Practice 5](#) is focused on reviewing the Hypergeometric Distribution.

Calculator Instructions

If you are using the TI-83/TI-84 series, there are probability functions for the binomial, Poisson, and geometric. Each has a pdf and a cdf (for example `binompdf` and `binomcdf`). These functions are located in 2nd DISTR. If you use, say, `binompdf(n, p)`, you will get the table of probabilities for 0, 1, 2, ..., n . If you use `binompdf(n, p)`, you will get the probability of x . If you use `binomcdf(n, p, x)`, you will get the cumulative probability ($P(X = 0) + P(X = 1) + P(X = 2) + \dots + P(X = n)$).

Assign Homework

Assign [Homework](#). Suggested homework: 1 - 17 odds, 23, 33 - 37 (Binomial and Poisson).

Continuous Random Variables: Teacher's Guide

This chapter is a good introduction to continuous types of probability distributions (the most famous of all is the normal). Two continuous distributions are covered – the uniform (or rectangular) and the exponential. For the uniform, probability is just the area of a rectangle. This distribution easily gets across the concept that probability is equal to area under a "curve" (a function). The exponential, which is used in industry and models decay, is a nice lead-in to the normal. The uniform and exponential distributions are also nice distributions to start with when you teach the Central Limit Theorem. It is interesting to note that the amount of money spent in one trip to the supermarket follows an exponential distribution. Several of our students discovered this idea when they chose data for their second project.

Compare Binomial v. Continuous Distribution

Begin this chapter by a comparison of a binomial (discrete) distribution and a continuous distribution. Using the normal for this comparison works well because the students are already familiar with it. The binomial graph has probability = height and the normal graph has probability = area. Tell the students that the discovery of probability = area in the continuous graph comes from calculus (which most of them have not studied). Draw the two graphs to make these ideas clear.

Introduce Uniform Distribution

Introduce the uniform distribution using the following example: The amount of time a student waits in line at the college cafeteria is uniformly distributed in the interval from 0 to 5 minutes (the students must wait in line from 0 to 5 minutes - each time in this interval is equally likely). Note: all the times cannot be listed. This is different from the discrete distributions.

Example:

Let X = the amount of time (in minutes) a student waits in line at the college cafeteria. The notation for the distribution is $X \sim U(a, b)$ where

$a = 0$ and $b = 5$. The function is $f(x) = \frac{1}{5}$ where $0 < x < 5$. The pattern is $f(x) = \frac{1}{b-a}$ where $a < x < b$.

In this example $a = 0$ and $b = 5$. The function $f(x)$ where $0 < x < 5$ graphs as a horizontal line segment.



Because
 $0 < x < 5$, the
 maximum area =
 $(5)(\frac{1}{5}) = 1$, the
 largest
 probability
 possible.

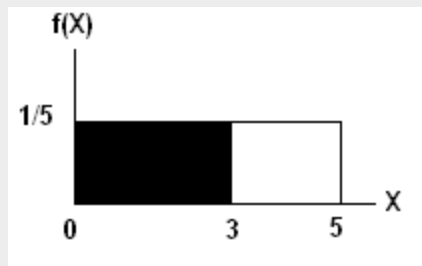
Example:

Exercise:

Problem:

Find the probability that a student must wait less than 3 minutes.
 Draw the picture and write the probability statement.

Solution:



Probability statement:
 $P(X < 3) = (3 - 0)(15) = 35$

.

The probability is the shaded area (the area of a rectangle with base $= b - a = 3 - 0 = 3$ and height $= \frac{1}{5}$). The probability a student must wait in the cafeteria line less than 3 minutes is $\frac{3}{5}$.

Example:

Exercise:

Problem: Find the average wait time.

Solution:

$$\mu = \frac{a+b}{2} = \frac{0+5}{2} = 2.5 \text{ minutes}$$

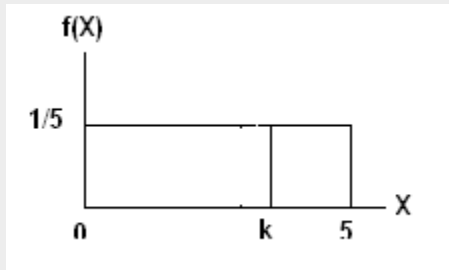
If the students take the time to draw the picture and write the probability statement, the problem becomes much easier.

Example:

Exercise:

Problem:

Find the 75th percentile of waiting times. A time is being asked for here. Percentiles often confuse students. They see "75th" and think they need to find a probability. Draw a picture and write a probability statement. Let k = the 75th percentile.

Solution:

- Probability statement: $P(X < k) = 0.75$
- Area: $(k - 0)\left(\frac{1}{5}\right) = 0.75 \quad k = 3.75$ minutes

75% of the students wait at most 3.75 minutes and 25% of the students wait at least 3.75 minutes.

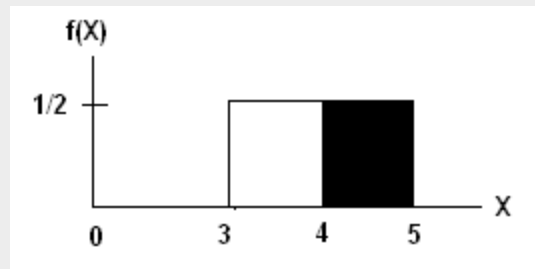
Example:**Exercise:****Problem:**

You can finish the uniform with a conditional. This reviews conditionals from [Continuous Random Variables](#). What is the probability that a student waits more than 4 minutes when he/she has already waited more than 3 minutes?

Solution:

$$\text{Algebraically: } P(X > 4 | X > 3) = \frac{P(X > 4 \text{ AND } X > 3)}{P(X > 3)} = \frac{P(X > 4)}{P(X > 3)}$$

Note: The students see it more clearly if you do the problem graphically. The lower value, a , changes from 0 to 3. The upper value stays the same ($b = 5$). The function changes to: $f(x) = \frac{1}{5-3} = \frac{1}{2}$



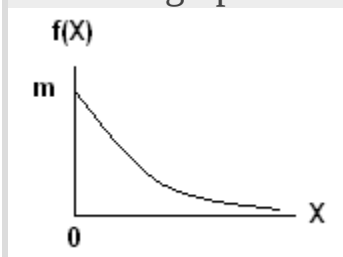
$$P(X > 4 | X > 3) = (\text{base})(\text{height}) = (5 - 4)\left(\frac{1}{2}\right) = \frac{1}{2}$$

Introduce the Change Example

The **exponential** distribution is generally concerned with how a quantity declines or decays. Examples include the life of a car battery, the life of a light bulb, the length of time business long distance telephone calls last, and the amount of change a person is carrying. You can introduce the exponential by using the change example. Ask everyone in your classroom to count their change and record it. Then have them calculate the mean and standard deviation and graph the histogram. The histogram should appear to be declining. Let X = the amount of change one person carries. Notation: $X \sim \text{Exp}(m)$ where m is the parameter that controls the amount of decline or decay; $m = \frac{1}{\mu}$ and $\mu = \frac{1}{m}$. Also, $\mu = \sigma$. (In the example, the calculated mean and standard deviation ought to be fairly close.)

Example:

The function is where $f(x) = me^{-mx}$ $m \geq 0$ AND $x \geq 0$. Find the probability that the amount of change one person has is less than \$.50. Draw the graph.



The right tail
extends
indefinitely.
There is no
upper limit in x .

The formula is $P(X < x) = 1 - e^{-mx}$ $P(X < .50) = \underline{\hspace{2cm}}$. The authors use technology to solve the probability problems. If you use the TI-83/84 calculator series, enter on the home-screen, $1 - e^{-m \cdot .50}$. Fill in the m with whatever the data produces ($m = \frac{1}{\mu}$; replace μ with the sample mean).

Ask the question, "Ninety percent of you have less than what amount?" and have them find the 90th percentile.

Draw the picture and let k = the 90th percentile. $P(X < k) = 0.90$. Solve the equation $1 - e^{-mk} = 0.90$ for k . On the home-screen of the TI-83/TI-84, enter $\frac{\ln(1-.90)}{(-m)}$.

Note: Have students fill in the blanks.

On average, a student would **expect** to have $\underline{\hspace{2cm}}$. The word "expect" implies the mean. Ten students together would **expect** to have $\underline{\hspace{2cm}}$. (the mean multiplied by 10)

Assign Practice

Assign the [Practice 1](#) and [Practice 2](#) in class to be done in groups.

Assign Homework

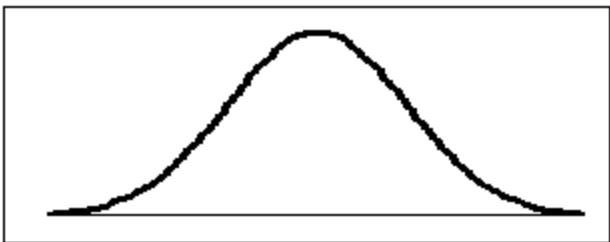
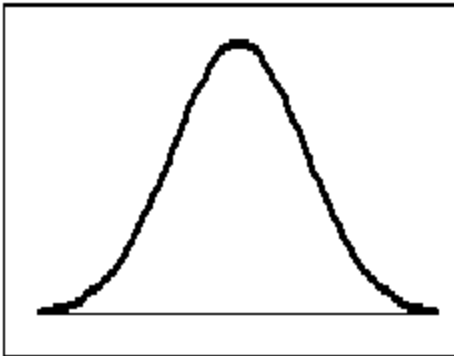
Assign [Homework](#). Suggested problems: 1 -13 odds, 15 - 20.

Normal Distribution: Teacher's Guide

A fair number of students are familiar with the "bell-shaped" curve. Stress that the normal is a continuous distribution like the uniform and exponential. However, the left and right tails extend indefinitely but come infinitely close to the x -axis. It is not necessary to show the probability distribution function for the normal (it is in the book) because there are normal probability tables and technology available for probability and percentile calculations.

Visualize the Data

Draw a picture of the normal graph and explain that it is symmetrical about the mean. The shape of the graph depends on the standard deviation. The smaller the standard deviation, the skinnier and taller the graph. A change in the mean shifts the graph to the right or left. The notation for the normal is $X \sim N(\mu, \sigma)$. Draw several normal curves (superimposed upon each other). Have students determine how the means and standard deviations are changing.



The Normal Distribution Notation

The standard normal distribution is of special interest. **Notation:** $Z \sim N(0,1)$ where Z = one z-score (the number of standard deviations a value is to the right or left of the mean). The mean is 0 and the variance (and standard

deviation) is 1. Any normal distribution can be standardized to the standard normal by the z-score formula: $z = \frac{\text{value} - \mu}{\sigma}$. Do an example showing the standardization. For $X \sim N(3, 2)$ and $Y \sim N(5, 6)$, the values $x = 4$ and $y = 8$ are each $\frac{1}{2}$ standard deviation to the right ($\frac{1}{2}\sigma$) of their respective means. Therefore, they both have a z-score of $\frac{1}{2}$.

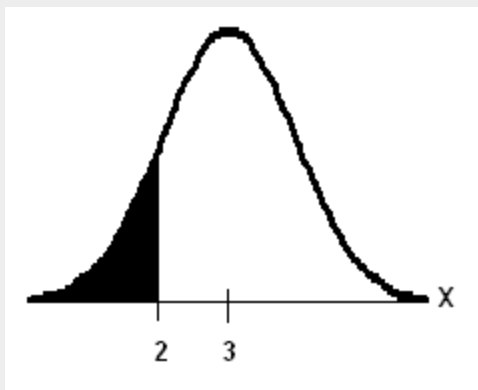
Example:

Do an example using the normal distribution and the standardization.

Exercise:**Problem:**

Several studies have shown that the amount of time people stand in line waiting for a bank teller is normally distributed. Suppose the mean waiting time is 3 minutes and the standard deviation is 1.5 minutes. Let X = the amount of time, in minutes, one person stands in line waiting for a teller. Notation: $X \sim N(3, 1.5)$

Find the probability that one person waits in line for a teller less than 2 minutes. Have students draw the picture and write a probability statement. The picture should have the x -axis.

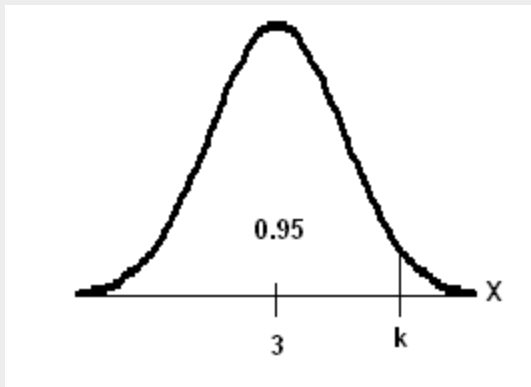
Solution:

Probability statement:
 $P(X < 2) = 0.2500$. If

you use the TI-83/84 series, the function

`normalcdf(0, 2,
3, 1.5)`

in 2nd DISTR. $k = 5.47$



$P(X < k) = 0.95$. If you use the TI83/84 series of articles, use the function `InvNorm(.95,3,1.5)`.
 $k = 5.47$

Note: The **normal approximation to the binomial is NOT included** in this text. With graphics calculators and computer software, it is easy to draw a binomial graph with a small n and then make n , say, 50. Students will see the graph approach the normal. The normal approximation states that if X follows a binomial distribution with number of trials equal to n and probability of success for any trial equal to p ($X \sim B(n, p)$), then by adding ± 0.5 to X , you get a new random variable Y (Y is either $X + 0.5$

or $X = 0.5$) and Y follows a normal distribution ($Y \sim N(np, npq)$). For the approximation to be a good one, you want $np > 5$, $nq > 5$, and $n > 20$.

Assign Practice

Assign the [Practice](#) in class to be done in groups.

Assign Homework

Assign [Homework](#). Suggested problems: 1 - 11 odds, 8, 10, 12 - 19.

Central Limit Theorem: Teacher's Guide

The Central Limit Theorem (CLT) is considered to be one of the most powerful theorems in all of statistics and probability. It states that if you draw samples of size n and average (or sum) them, you will get a distribution of averages (or sums) that follow a normal distribution.

Example:

Suppose μ and σ are the original mean and standard deviation of the population from which each sample of size n is drawn. Let \bar{X} = the random variable for the average of n samples. Let $\sum x$ = the random variable for the number of n samples

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

$$\sum x \sim N(n\mu, n\sigma)$$

The Dice Experiment

At the beginning of the chapter, there is a dice experiment. Together with the students, do the experiment. The example consists of rolling 10 times each, 1 die, 2 dice, 5 dice, and 10 dice and averaging the faces. Draw graphs (histograms are OK). This experiment, most of the time, shows that, as the number of dice increase, the graph looks more and more bell-shaped. Because the samples taken are usually small, you will not necessarily get a perfect bell-shaped curve. However, the students should get the idea.

Example:

Calculate Averages

It can be shown that the average amount of money one person spends on one trip to a particular supermarket is \$51. The averages follow an exponential distribution.

Exercise:

Problem:

Find the probability that the average of 40 samples is more than \$60.

Solution:

Let \bar{X} = the average amount of money that 40 people spend. Have the students draw the appropriate picture, labeling the x-axis with \bar{X} . The mean $\mu = 51$ and the standard deviation $\mu = 51$. If you are using the TI-83/84 series, use the function `normalcdf(60, 10^99, 51, 51/40)`.

The 75th percentile for the average amount spent by 40 people at the supermarket is \$56.44. This means that 75% of the people spend no more than \$56.44 and 25% spend no less than that amount.

This can be calculated by using the TI-83/84 function `InvNorm(.75, 51, 51/40)`.

Calculate Sums

You can also do examples for sums. We, the authors, do not do sums because of time (we are on a quarter system). Help the students to find the probability that the total (sum) amount of money spent by 10 people at the supermarket is less than \$500. Also, help them do a percentile problem.

Z-score Formulas

If you want to teach the z-score formulas for averages and sums, they are:

- $z = \frac{\text{value} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}$
- $z = \frac{\text{value} - n \cdot \mu}{\sqrt{n} \cdot \sigma}$

Assign Practice

Assign the [Practice](#) in class to be done in groups.

Assign Homework

Assign [Homework](#). Suggested homework: (averages) 1a - f, 3, 5, 9, 10, 11a - d, f, k, 13a-c,g-j, 16, 17, 19 - 23

Confidence Intervals: Teacher's Guide

Confidence intervals can be difficult for students. This chapter discusses confidence intervals for a single mean and for a single proportion. In this course, we do not deal with confidence intervals for two means or two proportions. For a single mean, confidence intervals are calculated when σ is known and when σ is not known (s is used as an estimate for σ).

Book notation:

- CL = confidence level
- EBM = error bound for a mean
- EBP = error bound for a proportion

The student-t distribution is introduced in this chapter beginning with a little history:

Note: William Gossett derived the t-distribution in 1908. He needed a method for dealing with small samples (less than 30) in his research on temperature at the Guinness Brewery. Legend has it that the name Student-t comes from the fact that Gossett wrote a paper about the t-distribution and signed the paper **Student** because he was too modest to use his own name.

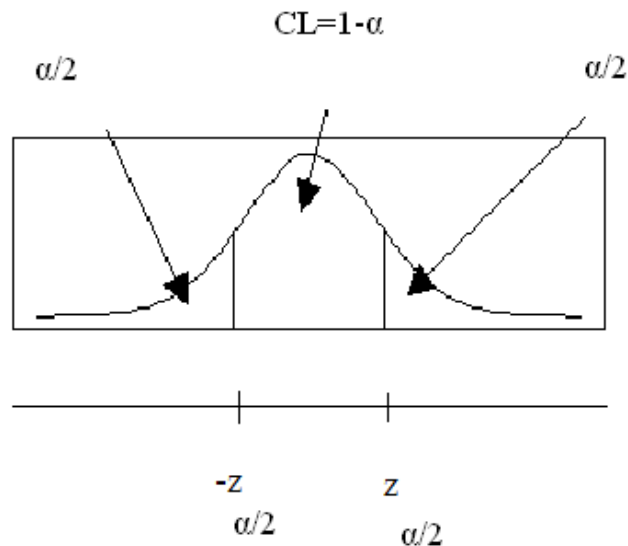
If you sample from a normal distribution in which σ is not known, replace σ with s , the sample standard deviation, and use the Student-t distribution. The shape of the curve depends on the parameter degrees of freedom (df). $df = n - 1$ where n is the sample size.

Note: t^{df} designates the distribution. We use T as the random variable. **Value** is an average.

The t-statistic (t-score)

$$t = \frac{\text{value} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}$$

The relationship between the confidence interval for a single mean (when σ and the confidence level can be shown in a picture as follows:



The $\frac{\alpha}{2}$ subscript indicates that the area to the right is $\frac{\alpha}{2}$.

Formulas for the error bounds:

- Single mean (known σ): $EBM = z_{\frac{\alpha}{2}} \cdot \left(\frac{\sigma}{\sqrt{n}} \right)$
- Single mean (unknown σ): $EBM = t_{\frac{\alpha}{2}} \cdot \left(\frac{s}{\sqrt{n}} \right)$
- Binomial proportion: $EBP = z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{p'q'}{n}}$ where $q = 1 - p'$

The confidence intervals have the form:

- Single mean (unknown or known σ): $(\bar{x} - EBM, \bar{x} + EBM)$
- Binomial proportion: $(p' - EBP, p' + EBP)$

Example:

Exercise:

Problem:

The number of calories in fast food is always of interest. A survey was taken from 7 fast food restaurants concerning the number of calories in 4 ounces of french fries. The data is 296, 329, 306, 324, 292, 310, 350. Construct a 95% confidence interval for the true average number of calories in a 4 ounce serving of french fries.

Solution:

You want a confidence interval for a single mean where σ is not known. If you use the TI-83/84 series, enter the data into a list and then use the function TInterval, data option. C-level is 95. The confidence interval is (296.4, 334.2). This function also calculates the sample mean (315.3) and sample standard deviation (20.4). TInterval is found in **STAT TESTS**.

If you want the students to use the formulas for a normal or for the Student-t confidence interval, you will need to use a table for the z-score or the t-score. The book does not have the tables but the Internet has several. Do a search on "z-score table" and "Student-t table."

First, you need to calculate the sample mean and the sample standard deviation.

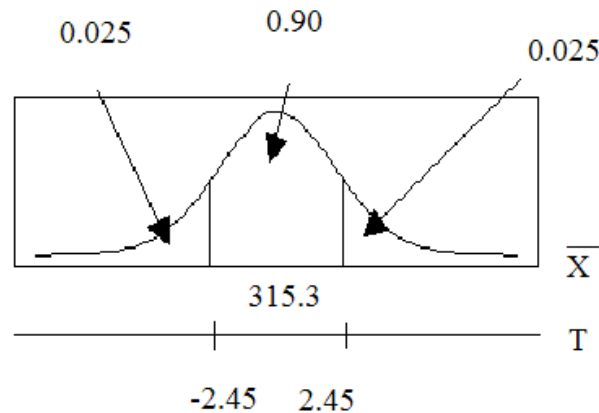
- $\bar{x} = 315.29$
- $s = 20.40$

The confidence interval has the pattern : $(\bar{x} - \text{EBM}, \bar{x} + \text{EBM})$

The error bound formula is : $\text{EBM} = t_{\frac{\alpha}{2}} \cdot \left(\frac{s}{\sqrt{n}} \right)$

CL = 0.95 so $\alpha = 0.05$. Therefore, $\frac{\alpha}{2} = 0.025$.

Using the Student-t table with $\text{df} = 7 - 1 = 6$, $t_{.025} = 2.45$.



$$EBM = t_{\frac{\alpha}{2}} \cdot \left(\frac{s}{\sqrt{n}} \right) = t_{0.025} \cdot \frac{20.40}{\sqrt{7}} = 2.45 \cdot \frac{20.40}{\sqrt{7}} = 18.89$$

The confidence interval is $(\bar{x} - EBM, \bar{x} + EBM) = (315.29 - 18.89, 315.29 + 18.89) = (296.4, 334.2)$

We are 95% confident that the true average number of calories in a 4 ounce serving of french fries is between 196.4 and 334.2 calories.

Example:

Exercise:

Problem:

At a local cabana club, 102 of the 450 families who are members have children who swam on the swim team in 1995. Construct an 80% confidence interval for the true proportion of families with children who swim on the swim team in any year.

Solution:

You want a confidence interval for a single proportion. If you use the TI-83/84 series, use the function 1-PropZinterval. $x = 102$, $n = 450$, $C\text{-level} = 80$. The confidence interval is (.2077, .2590)

If you want to use the formulas, first, you need to calculate the estimated proportion.

$$p' = \frac{x}{n} = \frac{102}{450} = 0.23$$

The confidence interval has the pattern $(p' - \text{EBP}, p' + \text{EBP})$.

The error bound formula is $\text{EBP} = z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{p'q'}{n}}$ where $q = 1 - p'$

CL = 0.80 so $\alpha = 0.20$. Therefore, $\frac{\alpha}{2} = 0.10$.

Using the normal table (find one on the Internet), $z_{.10} = 1.28$. (Remind students that 0.10 is the area to the right. The area to the left is 0.90.)

$$\text{EBP} = z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{p'q'}{n}} = z_{.10} \cdot \sqrt{\frac{p'q'}{n}} = z_{.10} \cdot \sqrt{\frac{.23 \cdot .77}{450}} = 1.28 \cdot \sqrt{\frac{.23 \cdot .77}{450}} = 0.03$$

The confidence interval is : $(p' - \text{EBP}, p' + \text{EBP}) = (0.23 - 0.03, 0.23 + 0.03) = (0.20, 0.26)$

We are 80% confident that the true proportion of families that have children on the swim team in any year is between 0.20 and 0.26.

Assign Practice

Assign the [Practice 1](#), [Practice 2](#), and [Practice 3](#) in class to be done in groups.

Assign Homework

Assign [Homework](#). Suggested homework: 1, 5, 9, 13, 15, 17, 21, 23, 24 - 31.

Hypothesis Testing: Single Mean and Single Proportion: Teacher's Guide

Hypothesis testing is done constantly in business, education, and medicine to name just a few areas. To perform a hypothesis test, you set up **two contradictory hypotheses** and use data to support one of them. Introduce the students to hypothesis testing by an example. Use a table to show the outcomes. Use H_o as the null hypothesis and H_a as the alternate hypothesis. Go over the language "reject H_o " and "do not reject H_a ".

Example:

Exercise:

Problem: H_o : John loves Marcia. H_a : John does not love Marcia.

- Type I error: Reject the null when the null is true.
 $P(\text{Type I error}) = \alpha$.
- Type II error: Do not reject the null when the null is false.
 $P(\text{Type II error}) = \beta$.
- Type I error: Marcia thinks John does not love her when he really does.
- Type II error: Marcia thinks John does love her when he does not.

Have the students try to write out the errors before you do. They may require a little prompting. Then have them state the possible consequences for the errors.

Conducting a Hypothesis Test

To perform the hypothesis test, sample data is gathered. The data typically favors one of the hypotheses (but not always). The test determines which hypothesis the data favors. If the data favors the null hypothesis, we "**do not reject**" the null hypothesis. If the data does not favor the null

hypothesis, we "**reject**" the null hypothesis. To not reject or to reject are decisions. After a decision is reached, an appropriate conclusion is made using complete sentences.

Sometimes the data favors neither hypothesis. In this case, we say the test is inconclusive.

A hypothesis test may be left-tailed, right-tailed or two-tailed. What the test is concerned with generally determines what type of test is being done.

Associated with the null hypothesis is a pre-conceived α . $\alpha = P(\text{Type I error})$. Students sometimes have a difficult time when there is no pre-conceived α . We use $\alpha = 0.05$ if there is none.

The data is used to calculate the p -value. The p -value is the probability that the information (data) will happen purely by chance when the null hypothesis is true. If we reject the null hypothesis, then we believe the information did not happen purely by chance with the current null hypothesis. Therefore, we believe that the null hypothesis is not true.

The decision (to reject or not reject) is based on whether $\alpha > p\text{-value}$ or $\alpha < p\text{-value}$.

The example in the book concerning Jeffrey, an eight-year old swimmer, is a good first example to do with the class. They can follow along in the book and then complete the problem that follows (bench press problem). By filling in the blanks, they are led through the steps of hypothesis testing.

In the beginning, the students have the most difficulty in determining which test to use (test of a single mean - normal or Student-t or test a binomial proportion) and the type (left-, right-, or two-tailed). We do several examples (usually we choose some homework problems) in class with the students. If a single mean Student-t is done, the assumption is that the population from which the data is taken is normal. In reality, this would have to be shown to be true.

[Here](#) is a series of solution sheets that can be copied and used by the students to do the hypothesis testing problems. A solution sheet makes it

clearer to the student what the steps to the tests are.

Go over the solution for "Fido's Fleas", a binomial proportion hypothesis testing problem written as a poem. The problem is at the end of the text portion of the chapter. The solution on a solution sheet follows the poem.

If you use the TI-83/84 series, there are functions to perform the different hypotheses tests. They can be found in STAT TESTS. Z-Test (normal test) does a test of a single mean when the population standard deviation is known; T-test (Student-t test) does a test of a single mean when the population standard deviation is not known; 1-PropZTest (normal test) does a test of a single proportion. The examples in the book contain TI-83/84 calculator instructions, in detail.

Assign Practice

Assign [Practice 1](#), [Practice 2](#), and [Practice 3](#) to be done collaboratively.

Assign Homework

Assign [Homework](#). Suggested problems: 1 - 15 odds, 19, 21, 25, 29, 31, 33, 34 - 44.

Assign Projects

There are two partner projects for this lesson: one uses an [article](#) and the other is a [word problem](#). Students create their own hypothesis testing problems and learn much from the process.

Hypothesis Testing: Two Population Means and Two Population Proportions: Teacher's Guide

The comparison of two groups is done constantly in business, medicine, and education, to name just a few areas. You can start this chapter by asking students if they have read anything on the Internet or seen on television any studies that involve two groups. Examples include diet versus hypnotism, Bufferin® with aspirin versus Tylenol®, Pepsi Cola® versus Coca Cola®, and Kellogg's Raisin Bran® versus Post Raisin Bran®. There are hundreds of examples on the Internet, in newspapers, and in magazines.

This chapter covers independent groups for two population means and two population proportions and matched or paired samples. **The module relies heavily on technology. Instructions for the TI-83/84 series of calculators are included for each example.** If you and your class are interested, the formulas for the test statistics are included in the text.

Doing problems 1 - 10 in the [Homework](#) helps the students to determine what kind of hypothesis test they should perform.

Example:

Matched or Paired Samples

A course is designed to increase mathematical comprehension. In order to evaluate the effectiveness of the course, students are given a test before and after the course. The sample data is:

Before Course	90	100	160	112	95	190	125
After Course	120	95	150	150	100	200	120

Example:

Two Proportions, Independent Groups

Suppose in the last local election, among 240 30-45 year olds, 45% voted and among 260 46-60 year olds, 50% voted. Does the data indicate that the proportion of 30-45 year olds who voted is less than the proportion of 46-60 year olds? Test at a 1% level of significance.

Firm A:

- $N_A = 20$
- $S_A = \$100$ $X_A = \$1500$

Firm B:

- $N_B = 22$
- $S_B = \$200$ $X_B = \$1900$

Test the claim that the average price of Firm A's laptop is no different from the average price of Firm B's laptop.

Calculator Instructions

If you use the TI83/84 series, the functions are located in STATS TESTS. The function for two proportions is 2-PropZTest, the function for two means is 2-SampTTest if the population standard deviations are not known and 2-SampZTest if the population standard deviations are known (highly unlikely). The function for matched pairs is T-test (the same test used for test of a single mean) because you combine two measurements for each object into a single set of "difference" data. For the function 2-SampTTest, answer "NO" to "Pooled."

Assign Practice

Have students do the [Practice 1](#) and [Practice 2](#) collaboratively in class. These practices are for two proportions and two means. For matched pairs,

you could have them do Example 10-7 in the text.

Assign Homework

Assign [Homework](#). Suggested homework problems: 1 - 10, 11, 13, 15, 17, 19, 23, 25, 31, 39 - 52.

Ch 11: The Chi-Square Distribution

This module is the complementary teacher's guide for the "The Chi-Square Distribution" chapter of the Collaborative Statistics collection (col10522) by Barbara Illowsky and Susan Dean.

This chapter is concerned with three chi-square applications: goodness-of-fit; independence; and single variance. We rely on technology to do the calculations, especially for goodness-of-fit and for independence. However, the first example in the chapter (the number of absences in the days of the week) has the student calculate the chi-square statistic in steps. The same could be done for the chi-square statistic in a test of independence.

The chi-square distribution generally is skewed to the right. There is a different chi-square curve for each df. When the df's are 90 or more, the chi-square distribution is a very good approximation to the normal. For the chi-square distribution, μ = the number of df's and σ = the square root of twice the number of df's.

Goodness-of-Fit Test

A goodness-of-fit hypothesis test is used to determine whether or not data "fit" a particular distribution.

Example:

In a past issue of the magazine *GEICO Direct*, there was an article concerning the percentage of teenage motor vehicle deaths and time of day. The following percentages were given from a sample.

Time of Day	Death Rate
12 a.m. to 3 a.m.	17%
3 a.m. to 6 a.m.	8%
6 a.m. to 9 a.m.	8%
9 a.m. to 12 noon	6%
12 noon to 3 p.m.	10%
3 p.m. to 6 p.m.	16%
6 p.m. to 9 p.m.	15%
9 p.m. to 12 a.m.	19%

Time of Day Percentage of Motor Vehicle Deaths

For the purpose of this example, suppose another sample of 100 produced the same percentages. We hypothesize that the data from this new sample fits a uniform distribution. The level of significance is 1% ($\alpha = 0.01$).

- H_o : The number of teenage motor vehicle deaths fits a uniform distribution.
- H_a : The number of teenage motor vehicle deaths does not fit a uniform distribution.

The distribution for the hypothesis test is χ^2_7

The table contains the observed percentages. For the sample of 100, the observed (O) numbers are 17, 8, 8, 6, 10, 16, 15 and 19. The expected (E) numbers are each 12.5 for a uniform distribution (100 divided by 8 cells). The chi-square test statistic is calculated using

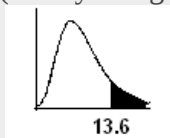
$$\sum_8 \frac{(O-E)^2}{E} = \frac{(17-12.5)^2}{12.5} + \frac{(8-12.5)^2}{12.5} + \frac{(8-12.5)^2}{12.5} + \frac{(6-12.5)^2}{12.5} + \frac{(10-12.5)^2}{12.5} + \frac{(16-12.5)^2}{12.5} + \frac{(15-12.5)^2}{12.5} + \frac{(19-12.5)^2}{12.5} = 13.6$$

If you are using the TI-84 series graphing calculators, ON SOME OF THEM there is a function in STAT TESTS called χ^2 GOF-Test that does the goodness-of-fit test. You first have to enter the observed numbers in one list (enter as whole numbers) and the expected numbers (uniform implies they are each 12.5) in a second list (enter 12.5 for each entry: 100 divided by 8 = 12.5). Then do the test by going to χ^2 GOF-Test.

If you are using the TI-83 series, enter the observed numbers in list1 and the expected numbers in list2 and in list3 (go to the list name), enter (list1-list2)^2/list2. Press enter. Add the values in list3 (this is the test statistic). Then go to 2nd DISTR χ^2 cdf. Enter the test statistic (13.6) and the upper value of the area (10^{-99}) and the degrees of freedom (7).

Probability Statement: $P(\chi^2 > 13.6) = 0.0588$

(Always a right-tailed test)



$$p\text{-value} = 0.0588$$

Since $\alpha < p\text{-value}$ ($0.01 < 0.0588$), we do not reject H_o .

We conclude that there is not sufficient evidence to reject the null hypothesis. It appears that the number of teenage motor vehicle deaths fits a uniform distribution. It does not matter what time of the day or night it is. Teenagers die from motor vehicle accidents equally at any time of the day or night. However, if the level of significance were 10%, we would reject the null hypothesis and conclude that the distribution of deaths does not fit a uniform distribution.

A **test of independence** compares two factors to determine if they are independent (i.e. one factor does not affect the happening of a second factor).

Example:

The following table shows a random sample of 100 hikers and the area of hiking preferred.

Gender	The Coastline	Near Lakes and Streams	On Mountain Peaks
Female	18	16	11
Male	16	25	14

Hiking Preference Area The two factors are gender and preferred hiking area.

- H_o : Gender and preferred hiking area are independent.
- H_a : Gender and preferred hiking area are not independent

The distribution for the hypothesis test is χ^2_2 .

The df's are equal to: $(\text{rows} - 1)(\text{columns} - 1) = (2 - 1)(3 - 1) = 2$

The chi-square statistic is calculated using $\sum_{(2-3)} \frac{(O-E)^2}{E}$

Each expected (E) value is calculated using $\frac{(\text{row total})(\text{column total})}{\text{total surveyed}}$

The first expected value (female, the coastline) is $\frac{45 \cdot 34}{100} = 15.3$

The expected values are: 15.3, 18.45, 11.25, 18.7, 22.55, 13.75

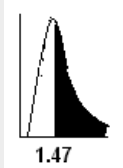
The chi-square statistic is:

$$\sum_{(2-3)} \frac{(O-E)^2}{E} = \frac{(18-15.3)^2}{15.3} + \frac{(16-18.45)^2}{18.45} + \frac{(11-11.25)^2}{11.25} + \frac{(16-18.7)^2}{18.7} + \frac{(25-22.55)^2}{22.55} + \frac{(14-13.75)^2}{13.75} = 1.47$$

Calculator Instructions

The TI-83/84 series have the function χ^2 -Test in STAT TESTS to perform this test. First, you have to enter the observed values in the table into a matrix by using 2nd MATRIX and EDIT [A]. Enter the values and go to χ^2 -Test. Matrix [B] is calculated automatically when you run the test.

Probability Statement: $p\text{-value} = 0.4800$ (A right-tailed test)



$$p\text{-value} = 0.4800$$

Since α is less than 0.05, we do not reject the null.

There is not sufficient evidence to conclude that gender and hiking preference are not independent.

Sometimes you might be interested in how something varies. A **test of a single variance** is the type of hypothesis test you could run in order to determine variability.

Example:

Exercise:

Problem:

A vending machine company which produces coffee vending machines claims that its machine pours an 8 ounce cup of coffee, on the average, with a standard deviation of 0.3 ounces. A college that uses the vending machines claims that the standard deviation is more than 0.3 ounces causing the coffee to spill out of a cup. The college sampled 30 cups of coffee and found that the standard deviation was 1 ounce. At the 1% level of significance, test the claim made by the vending machine company.

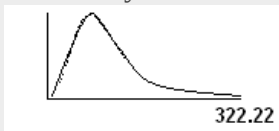
Solution:

$$H_o:\sigma^2 = (0.3)^2 \quad H_a:\sigma^2 > (0.3)^2$$

The distribution for the hypothesis test is χ^2_{29} where $df = 30 - 1 = 29$.

$$\text{The test statistic } x^2 = \frac{(n-1) \cdot s^2}{\sigma^2} = \frac{(30-1) \cdot 1^2}{0.3^2} = 322.22$$

Probability Statement: $P(x^2 > 322.22) = 0$



$$p\text{-value} = 0$$

Since $\alpha > p\text{-value}$ ($0.01 > 0$), reject H_o .

There is sufficient evidence to conclude that the standard deviation is more than 0.3 ounces of coffee. The vending machine company needs to adjust their machines to prevent spillage.

Assign Practice

Have the students do the [Practice 1](#), [Practice 2](#), and [Practice 3](#) in class collaboratively.

Assign Homework

Assign [Homework](#). Suggested homework: 3, 5, 7 (GOF), 9, 13, 15 (Test of Indep.), 17, 19, 23 (Variance), 24 - 37 (General)

Ch 12: Linear Regression and Correlation

This module is the complementary teacher's guide for the "Linear Regression and Correlation" chapter of the Collaborative Statistics collection (col10522) by Barbara Illowsky and Susan Dean.

Entire courses are given on linear regression and correlation. This chapter serves as an introduction to the topics.

It helps to review the equation of a line. We use a for the y -intercept and b for the slope. The line has the form: $y = a + bx$

Example:

Exercise:

Problem:

Have the students plot a line by eye using the following data. The independent variable x represents the size of a color television screen in inches at Anderson's and y represents the sales price in dollars.

x	9	20	27	31	35	40	60
y	147	197	297	447	1177	2177	2497

Ask them what they got for the slope and for the y -intercept. Make comparisons. This exercise should point out how difficult it is to get an accurate line of best fit and how many lines "seem" to fit the data. (This data is taken from the exercises.)

Solution:

For the data above, use either a calculator or a computer and calculate the least squares or best fit line. Look at the scatter plot first. Ask the students if their "by eye" line looks like the calculated one. Explain the correlation coefficient and then check if the correlation coefficient is significant by comparing it to the correct entry in 95% CRITICAL VALUES OF THE SAMPLE CORRELATION COEFFICIENT Table at the end of the reading.

If you use the TI-83/84 series, enter the data into two lists first. Then plot the data points on the calculator. First set up the stat plot (2nd STAT PLOT). Then press ZOOM 9 to see the plot. To do the linear regression, go to the LinReg ($a + bx$) function in STAT CALC. Enter the lists. At this time, you could also enter a y-variable after the lists (after you enter the lists, enter a comma and then press VARS Y-VARS Function Y1). Press ENTER to see the linear regression. When you press GRAPH, the line will plot.

Line of best fit: $\hat{y} = -745.2420 + 54.7557x$.

Explain "predicting" (or forecasting) and have them predict the sales price of a 45 inch screen color TV. Have them predict the cost for a mini 5 inch color TV. (The answer is negative.) Discuss that the line is only valid from the lowest to the highest x - values.

Example:

Exercise:

Problem:

Have the students follow the "outlier" example in the text and (just once!) do the calculations for finding an outlier. Have them fill in the table below.

x	y	$y - \text{yhat}$	$ y - \text{yhat} $	$(y - \text{yhat})^2$

Find: $\sum_7 (|y - \text{yhat}|)^2 = \text{SSE}$

Find $s = \sqrt{\frac{\text{SSE}}{n-2}}$

n = the total number of data values (7 for this problem)

s is the standard deviation of the $|y - \text{yhat}|$ values

Multiply s by 1.9: $(1.9)(s) = \underline{\hspace{2cm}}$

Compare each $|y - \text{yhat}|$ to $(1.9)(s)$.

If any $|y - \text{yhat}|$ is at least $(1.9)(s)$, then the corresponding point is an outlier. (None of the points is an outlier.)

Assign Practice

Have the students do the [Practice](#) collaboratively in class.

Assign Homework

Assign [Homework](#). Suggested homework: 1, 3, 5, 9, 13, 15 (a - f only if you use the calculator), 21 - 25.

Ch 13: F Distribution and ANOVA

This module is the complementary teacher's guide for the "F Distribution and ANOVA" chapter of the Collaborative Statistics collection (col10522) by Barbara Illowsky and Susan Dean.

Note: The F distribution is named after Ronald Fisher. Fisher is one of the most respected statisticians of all time. He did a lot of statistical work in biology and genetics and became chair of genetics at Cambridge University in England in 1949. In 1952, he was awarded knighthood.

This section is a **very brief overview** of the F distribution and two of its applications - One Way Analysis of Variance (ANOVA) and test of two variances. There are college courses which deal exclusively with these topics. ANOVA, particularly, is used regularly in industry.

Explanation of Sum of Squares, Mean Square, and the F ratio for ANOVA

- k = the number of different groups
- n_j = the size of the j th group
- s_j = the sum of the values in the j th group
- N = the total number of all the values combined
- Total sample size: $\sum n_j$
- x = one value: $\sum x = \sum s_j$
- Sum of squares of all values from every group combined: $\sum x^2$
- Between group variability: $SS_{\text{total}} = \sum x^2 - \frac{(\sum x)^2}{N}$
- Total sum of squares: $\sum x^2 - \frac{(\sum x)^2}{N}$
- Explained variation- sum of squares representing variation among the different samples $SS_{\text{between}} = \sum \left[\frac{(s_j)^2}{n_j} \right] - \frac{(\sum s_j)^2}{N}$
- Unexplained variation- sum of squares representing variation within samples due to chance: $SS_{\text{within}} = SS_{\text{total}} - SS_{\text{between}}$
- df's for different groups (df's for the numerator): $df_{\text{between}} = k - 1$

- Equation for errors within samples (df's for the denominator):

$$df_{\text{within}} = N - k$$
- Mean square (variance estimate) explained by the different groups:

$$MS_{\text{between}} = \frac{SS_{\text{between}}}{df_{\text{between}}}$$
- Mean square (variance estimate) that is due to chance (unexplained):

$$MS_{\text{within}} = \frac{SS_{\text{within}}}{df_{\text{within}}}$$
- F ratio or F statistic of two estimates of variance: $F = \frac{MS_{\text{between}}}{MS_{\text{within}}}$

Note: The above calculations were done with groups of different sizes. If the groups are the same size, the calculations simplify somewhat and the F ratio can be written as:

Equation:

F Ratio Formula

$$F = \frac{n \cdot (s_{-x})^2}{(s_{\text{pooled}})^2}$$

where...

- $(s_{-x})^2$ = the variance of the sample means
- n = the sample size of each group
- $(s_{\text{pooled}})^2$ = the mean of the sample variances (pooled variance)
- $df_{\text{numerator}} = k - 1$
- $df_{\text{denominator}} = k(n - 1) = N - k$

These calculations are easily done with a graphing calculator or a computer program. We present the information in the chapter assuming some kind of technology will be used.

For ANOVA, the samples must come from normally distributed populations with the same variance, and the samples must be independent. The ANOVA test is right-tailed.

In a test of two variances, the samples must come from normal populations and must be independent of each other.

Exercise:

Problem:
(One-Way ANOVA)

Three different diet plans are to be tested for average weight loss. For each diet plan, 4 dieters are selected and their weight loss (in pounds) in one month's time is recorded.

Plan 1	Plan 2	Plan 3
5	3.5	8
4.5	7	4
4	6	3.5
3	4	4.5

Is the average weight loss the same for each plan? Conduct an ANOVA test with a 1% level of significance.

Solution:

Let μ_1 , μ_2 , and μ_3 be the population means for the three diet plans.

- $H_o: \mu_1 = \mu_2 = \mu_3$
- H_a : Not all pairs of means are equal.
- $df_{\text{numerator}} = 3 - 1 = 2$
- $df_{\text{denominator}} = 12 - 3 = 9$

The distribution for the test is $F_{2,9}$

Using a calculator or computer, the test statistic is $F = 0.47$. The notation used for the F statistic may also be F' or $F_{2,9}$ (like the distribution). The TI-83/84 series has the function ANOVA in STAT TESTS. Enter the lists of data separated by commas.

If you use the formulas for groups of the same size, the calculations are as follows:

Sample means are 4.13, 5.13, and 5, respectively. Sample standard deviations are 0.8539, 1.6250, and 2.0412, respectively.

$(s_{-x})^2 = 0.2956$	The variance of the sample means
$(s_{\text{pooled}})^2 = 2.5416$	The mean of the sample variances
$n = 4$	The sample size of each group

Equation:

$$F = \frac{4 \cdot 0.2956}{2.5416}$$

Probability Statement: $P(F > 0.47) = 0.6395$



$$p\text{-value} \\ = 0.6395$$

Since $\alpha < p\text{-value}$, do not reject H_o .

There is not sufficient evidence to conclude that the three diet plans are different. It appears that the three diet plans work equally well. The average weight loss is the same for all three plans.

Exercise:

Problem:

(Test of Two Variances):

Machine A makes a box and machine B makes a lid. For the lid to fit the box correctly, the variances should be nearly the same. There is a suspicion that the variance of the box is greater than the variance of the lid. The following data was collected.

	Machine A (Box)	Machine B (Lid)
Number of Parts	9	11
Variance	150	45

Are the machines working properly? Test at a 5% level of significance.

Solution:

Let σ_A^2 and σ_B^2 be the population variances for machine A and machine B, respectively

- $H_o: \sigma_{A^2} = \sigma_{B^2}$
- $H_a: \sigma_{A^2} > \sigma_{B^2}$
- $n_A = 9$
- $n_B = 11$
- $df_{\text{numerator}} = 9 - 1$
- $df_{\text{denominator}} = 11 - 1 = 10$

The distribution for the hypothesis test is $F_{8,10}$

If you are using the TI-83/84 calculators, use the function 2-SAMPFTest for the test.

Using the formulas,

The test statistic is $F = \frac{[\frac{(s_A)^2}{(\sigma_A)^2}]}{[\frac{(s_B)^2}{(\sigma_B)^2}]} =$

$$\frac{(s_A)^2}{(s_B)^2} = \frac{150}{45} = 3.33$$

Since $\alpha > p\text{-value}$, reject the null hypothesis.

There is sufficient evidence to conclude that the box and lid do not fit each other. The variance of the box is larger.

Assign Practice

Have the students work collaboratively to complete the [Practice](#).

Assign Homework

Assign [Homework](#). Suggested homework: 1, 3, 4, 5.